Multimedia Clip Type: Quality of Perception Impact on Users With and Without Hearing Loss.

S. R. Gulliver and G. Ghinea

Department of Information Systems and Computing,
Brunel University,
Uxbridge, Middlesex,
UB8 3PH, UK
(Stephen.Gulliver / George.Ghinea)@brunel.ac.uk

ABSTRACT

This paper investigates how variance in multimedia video clip type affects quality of perception (QoP) for users with and without hearing loss. QoP encompasses not only a user's satisfaction with the quality of a multimedia presentation (subjective), but also his or her ability to analyse, synthesise and assimilate its' informational content (objective). Results show that clip type has a significant impact on the level of deaf information assimilation. Results suggest that certain video content aids deaf information assimilation, for example: those with less textual content. However, it was found that audio / captioned information does not significantly impact user QoP, when Video-textual (VT) information was found to have a significant effect on both hearing and deaf QoP. A positive correlation was found between predicted level of information assimilation and level of enjoyment, independent of hearing level or clip type.

1. Introduction

There are two main approaches to the assessment of digital multimedia quality: *objective* testing (via appropriately defined quality metrics) and *subjective* testing (which uses human viewers). Traditionally, the quality of distributed multimedia video has been defined using objective QoS (Quality of Service) technical parameters, such as jitter, delay, as well as loss and error rates. Although measurable, such parameters often disregard variance in user perception (Apteker et al, 1995). Extensive study of multimedia perception has been made in the fields of educational psychology and HCI. However, these studies largely concentrate on either the ergonomical and presentational aspects of the display, such that user perception is improved (Ghinea and Thomas, 1999), or the effect that objective parameters (such as jitter, packet loss, delay and frame rate) have on video quality metrics (Bouch and Sasse, 2000) (Kawalek, 1995). This paper investigates how variance in multimedia video clip type, independent of QoS parameter variance, affects QoP (Quality of Perception) for users with and without hearing loss.

2. Problem Defining Quality

Quality is a term that means different things to different people (Watson and Sasse, 2000), yet it is clear that multimedia quality is not a clear single monotone dimension. Multimedia video contains multiple media, both time-dependent and time-independent in nature (Apteker et al, 1995). Multiple variables, both objective and subjective, are often required to effectively specify the quality of each media. For example, speech quality can be determined using variables such as intelligibility, loudness, naturalness, listening effort and pleasantness of tone (Watson and Sasse, 2000). Due to the multi-dimensional complexity of multimedia, it is therefore impossible to define the quality of any multimedia video by simply measuring any one specific influencing factor.

To date, there has been a common assumption in the distributed networking community that quality issues, relating to distributed digital video, will be resolved through objective solutions, for example: increased bandwidth allocation (Jayant, 1990), or constant monitoring of perceptual metrics. Unfortunately, due to the multi-dimensional nature of multimedia, it seems impossible to rely purely on objective factors alone when defining multimedia quality.

Whilst investigating the relative impact of degradation on subjective quality ratings for interactive speech, transmitted over packet-switched networks, Watson and Sasse found that dissatisfaction with the quality was rarely due to objective factors such as packet loss or jitter (Watson and Sasse, 2000). Their results showed that the effects of subjective factors such as volume differences, echo and the use of a bad microphone, were rated as more influential to the perceived level of quality than the effect of packet loss.

As multimedia applications are produced for the enjoyment and/or education of human viewers, the viewers' opinion of the presentation quality is surely important to any quality definition.

3. Defining Quality Using Perception (QoP)

Ghinea and Thomas first introduced the notion of QoP, as they realised that objective factors alone were incapable of defining the perceived quality of distributed multimedia video (Ghinea and Thomas, 1998). QoP uses both

subjective 'level of enjoyment' (LOE) and objective level of 'information assimilation' (IA) to determine the overall perceived level of video quality.

3.1 Defining Information Assimilation (IA)

After showing a participant a short video clip, the participant is asked a number of IA feedback questions. These questions are used to determine the type and level of information assimilated by each user. IA questions have a definite answer, for example: (from the rugby video-clip used in this experiment) "What was the score of the match, before the try?" As this question has a single and definite answer, it is possible to determine whether the participant answered this question correctly or not. IA questions are designed so they can only be answered if certain information is assimilated from a specific information source (for example, the number of lions in the documentary is specific to the video). Distribution and use of questions is dependent on the video characteristic weightings (see Table 1). QoP information sources, used in this study, include:

- C/A : Information which is presented in both the Audio stream and (transcribed) in the Caption window (21 questions in total used in this study).
- C : Information from Captions contained in the video window, for example: the newsreader's name (4 questions in total used in this study).
- T : Textual information contained in the video window, yet not contained in any captions, for example: the number on a rugby shirt (12 questions in total used in this study).
- D : Dynamic information contained in the video window, for example: whether an actor exited to the left or right of the screen (12 questions in total used in this study).
- V : Other information relating to the Video window, for example: the colour of the presenters shirt, in the documentary clip (40 questions in total used in this study).

Video characteristic weightings, the importance that various information sources have within a specific video, are defined using participant feedback. Original definition of characteristic weightings failed to consider the use of additional captions. To ensure consistent definition with captions, C and T are defined together with weighting 'Video Text'. The multimedia video clips used in this experiment were varied in nature and chosen to cover a broad spectrum of possible video types (see Table 1).

Video Title	Dynamic (D)	Audio (C/A)	Video (V)	Video Text (VT)
Space (Action Sequence)				
Band				
Snooker				
Cooking				
News				
Rugby				
Pop Music				
Documentary				
Animated Movie				
Weather Forecast				

Table 1: Video characteristic weightings (white – no importance (0),grey – some importance (1), black – important (2)).

3.2 Defining Level of Enjoyment (LOE)

For each video clip, all participants were asked on a scale (0-10 for this study) how much they enjoyed the video clip. This information is used to determine whether ability to assimilate information has any relation to personal preference and level of enjoyment.

3.3 Defining Self–Predicted Level of Information Assimilation (PIA)

For each video clip, all participants were asked on a scale (0-10 for this study), how much information they thought they had assimilated. By asking this question we were able to judge what proportion of the available information the participant thought they assimilated. Although not essential to QoP definition, PIA was added to allow us to analyse how much information the participants' perceived to be available and how accurately they judged themselves to be at QoP IA.

4. Experimental Method

In order to ensure consistent experimental conditions for all participants, a series of measures were rigorously follow during our study.

- Constant frame rate (25 frames per second), stereo sound (16-bit sampled at 44.1 KHz) and colour depth (24-bit) were used for all video clips.
- Deaf participants were required to provide full details about their type of hearing loss and when it occurred. Self-defined deafness level was then used to prevent lengthy hearing tests or user embarrassment.
- Experimental conditions (level of lighting, seating angle {90° to screen} and distance from screen {60-80cm}) and equipment (233MHz MMX, with 16 bit 3D sound and a Trident Cyber 9397 3D Video Graphics card) were considered for all interviews.
- To reduce background noise and ensure a consistent audio level for all participants, headphones were used.
- If a participant normally required spectacles, then they were told to use them for the duration of the experiment. A short vision check was made for all participants.

46 participants (16 deaf {D} and 30 hearing {H}) were divided into groups 1, 2, 3 and 4. Groups 1 and 3 were of equal size and were made up of deaf participants. Groups 2 and 4 were also of equal size and were made up of hearing participants. Group 1 and 3 both consisted of a 4 / 4 split of pre-lingual profoundly deaf BSL (British sign language) users and post-lingual mild and moderately deaf non-BSL users. Although these two deaf groups alone are not representative of the total deaf community (Royal National Institute for the Deaf, 2002), they represent the range of deafness possible. Participants in all groups were taken from a range of different ages backgrounds. Division of participants into four groups, with different combinations of clip-order and captioning, was applied to minimise order-effects and participant frustration with captions. For detailed information concerning which captions were shown to which groups, see Table 2.

Video Clip	Group 1	Group 2	Group 3	Group 4
	(D)	(H)	(D)	(H)
Space (Action Movie)	Off	On	On	Off
Band (Jazz Band)	On	Off	Off	On
Snooker	Off	On	On	Off
Cooking	On	Off	Off	On
News (Local news)	Off	On	On	Off
Rugby (World Cup)	On	Off	Off	On
Pop (MTV style music)	Off	On	On	Off
Doc (Lion Documentary)	On	Off	Off	On
Animation (101 Dalmatians)	Off	On	On	Off
Forecast (National Weather)	On	Off	Off	On

Table 2: Use of Captions for Groups 1 to 4 (ON – Captions; OFF – No Captions)

It was explained to all participants that the experiment involved watching 10 short video clips and that after each video clip, they would be required to stop and answer a number of questions about the clip that had just been presented to them. It was made clear to each participant their intelligence was not being tested and that they should not be concerned if they were unable to answer any of the questions. After the user had been shown a video clip, QoP questions were asked and all responses were noted. Sign language translation was provided for those participants' using British Sign Language (BSL) as their first language.

5. Results

5.1 Effect of Video Clip Type on Information Assimilation (IA)

The level of hearing IA in this study varies between 59% and 72.5% depending on the type of video clip (see Figure 1). The level of deaf IA varies between 26.9% and 60.1%, depending on the type of video clip (see Figure 1). Multiple analysis of variance showed that video clip type has a significant affect on deaf IA $\{F(9,16) = 7.388, p<0.001\}$. No significant affect was measured for hearing IA.

Previous work has shown that level of hearing causes a significant difference between deaf and hearing information assimilation (Gulliver and Ghinea 2000). To analyse the affect that characteristic weightings have on user IA, univariant analysis of variance was performed for all videos shown during this study. Dynamic (D), caption / audio (C/A), video (V) and video text (VT) characteristics were used as fixed variables and IA, %PIA and %LOE were treated separately as dependent variables for hearing and deaf participants. This allows us to identify what characteristic weightings significantly impact QoP for users with and without hearing loss.

- Deaf IA: Video text (VT) $\{F(2,160) = 10.902, p<0.001\}$ was found to have a significant impact on deaf IA.
- Deaf PIA (%): Video text (VT)$\{F(2,160) = 5.551, p = 0.005\}$ was found to have a significant impact on deaf % PIA.
- Deaf LOE (%): Video Text (VT) $\{F(2,160) = 6.963, p= 0.001\}$ was found to have a significant impact on overall deaf % LOE.
- Hearing IA: No specific characteristic was found to have a significant impact on hearing IA.
- Hearing PIA: No specific characteristic was found to have a significant impact on hearing % PIA.

- Hearing LOE: Audio (C/A) {F(2,300) = 12.927, p < 0.001} and Video text (VT) {F(2,300) = 11.703, p < 0.001} were found to have a significant impact on hearing user % LOE.

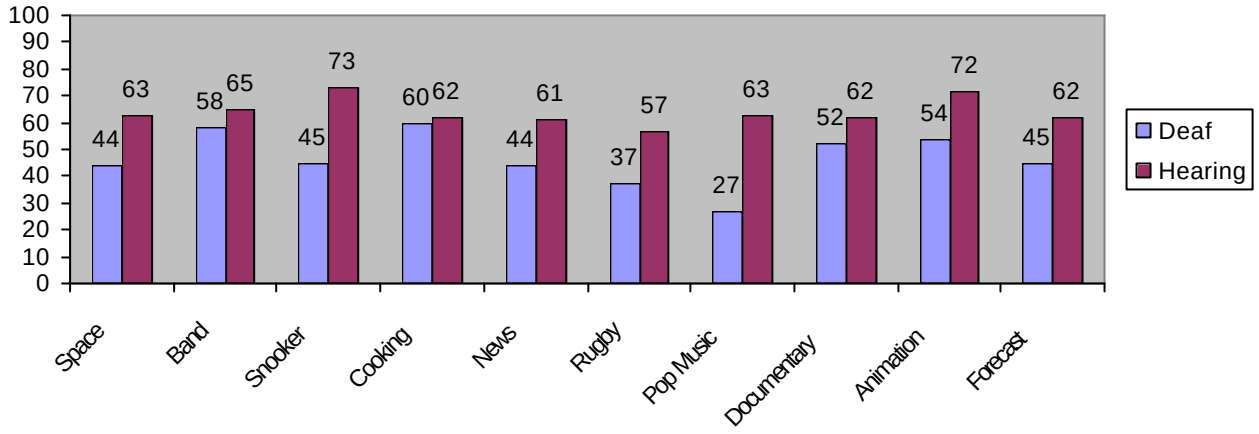

Figure 1: Percentage of Information Assimilation for Deaf and Hearing groups.

As the distribution of IA questions is dependent on the video characteristic weightings, consistent assimilation across all video clips is dependent on a participant assimilating the same level of information from all information sources. As different characteristic weightings have a significant impact on user IA, consistent overall QoP cannot be guaranteed. A breakdown of deaf and hearing QoP IA, using the defined information sources (C/A, C, T, D, V), showed that the percentage of deaf correctly assimilated information varies considerably between different information sources (see Figure 2). It appears that significant variation in deaf information assimilation, mainly video textual information, causes clip type to having a significant impact on deaf QoP IA.

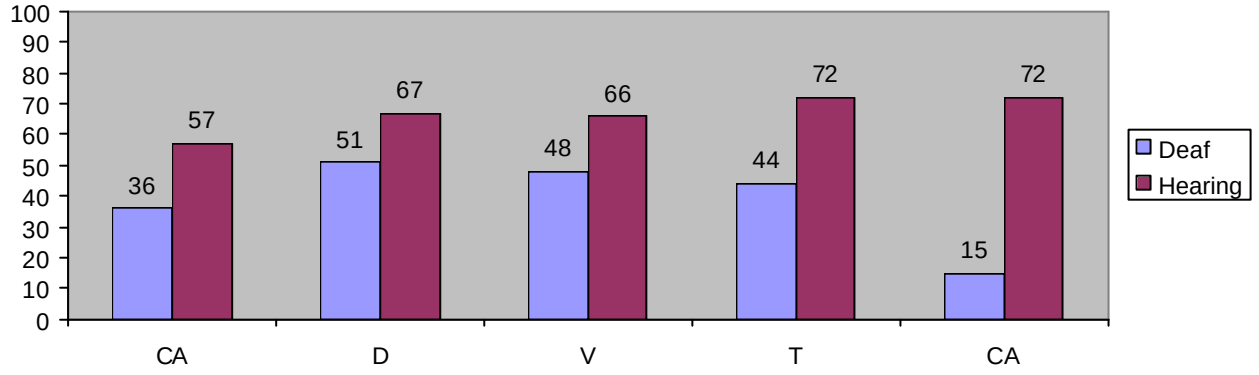

Figure 2: A detailed breakdown of the deaf / hearing IA (%).

5.2 Effect of Video Clip Type on Experienced Level of Enjoyment (LOE)

Multiple analysis of variance showed that video clip type has a significant affect on both hearing {F(9,30) = 38.266, p<0.001} and deaf {F(9,30) = 9.599, p<0.001} level of enjoyment. Therefore, video clip type has a significant impact on user perceived level of enjoyment, the second essential component (beside information analysis, synthesis and assimilation) of QoP, independent of user level of hearing. This shows that clip type affects a hearing participant's level of enjoyment, even though the level of information assimilated is not objectively affected.

5.3 Effect of Video Clip Type on Predicted Information Assimilation (PIA)

Multiple analysis of variance showed that video clip type has a significant affect on both hearing {F(9,30) = 4.993, p<0.001} and deaf {F(9,16) = 4.228, p<0.001} predicted level of information assimilation. During the experiment, deaf users often appeared to preconceive their own abilities for certain video clips. For example: profoundly deaf participants appeared to dismiss the 'Pop Music' clip as a waste of time, ignoring the fact that information required to answer feedback questions was also displayed in captions. Despite this general dismissal, no significant correlation was measured between deaf level of information assimilation and deaf predicted level of information assimilation.

Clip type significantly affects a participant's self-predicted level of information assimilation, although no objective increase in IA is measured (see Figure 3). A positive correlation was found between predicted level of

information assimilation and level of enjoyment, for both deaf {N=20, CC=.327, p=.047} and hearing {N=20, CC=.327, p=.026} participants.

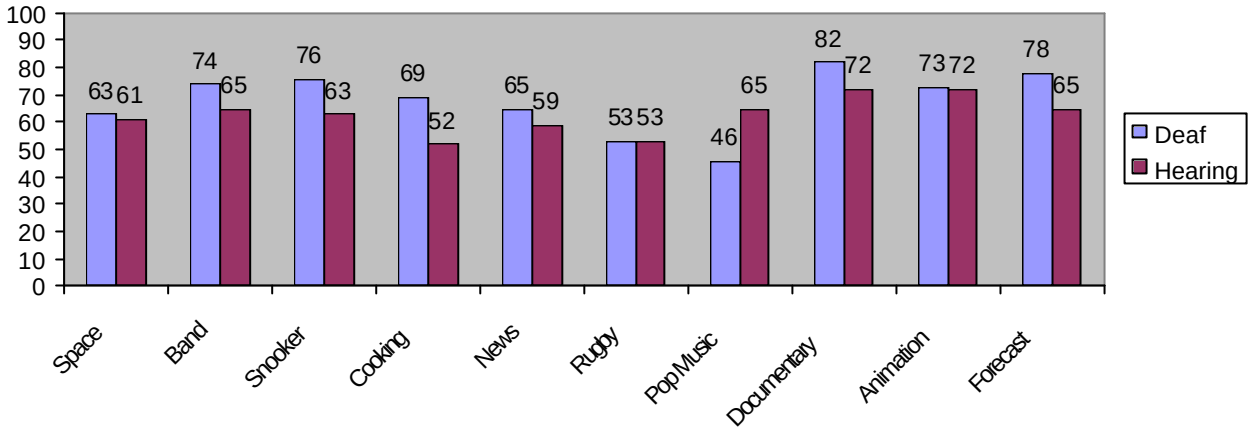

Figure 3: Predicted level of Information Assimilation (%), for Deaf and Hearing groups.

6. Conclusion

This paper investigates how variance in multimedia video clip type affects Quality of Perception (QoP) for users with and without hearing loss.

Results showed that clip type has a significant impact on deaf level of information assimilation. No significant affect was measured for hearing IA, which suggests that clip type has a more significant impact on deaf QoP IA than it does on hearing QoP IA. Results suggest that the content of certain video clips are easier for deaf participants to assimilate than others, for example: those with less textual content. Although this may seem common sense, it is interesting that audio / captioned information does not significantly impact user QoP, yet Video-textual (VT) information was found to have a significant effect on both hearing and deaf QoP.

Hearing participants predicted themselves as assimilating significantly more from certain video clips, although no significant difference in objective IA was measured. Type of video clip did, however, significantly impact hearing predicted level of information assimilation and level of enjoyment (the subjective factors). A positive correlation was found between predicted level of information assimilation and level of enjoyment, independent of hearing level or deafness type.

In concluding, we identify that multimedia clip type can has a significant impact on user QoP, independent of hearing level. The fact that multimedia video clip type, a measurement relatively independent of QoS variation, significantly impacts user QoP has serious implications on current approaches to video quality definition. Although further work is required to identify the impact that both objective and subjective parameters have on user QoP, it seems clear that future methods of video quality definition must consider both objective and subjective testing if video quality definition is not going to disregard the users' view of what defines quality.

REFERENCES

Apteker R. T., Fisher J. A., Kisimov V. S., and Neishlos H. (1995). Video Acceptability and Frame Rate, *IEEE Multimedia*, pp. 32-40, 2(3).

Bouch A., Sasse M. A., and DeMeer H. (2000). Of Packets and People: A User-Centred Approach to Quality of Service. *Proceedings of IWQoS 2000*, Pittsburgh, PA, June 5-8, pp. 189-197.

Ghinea G., and Thomas J. P. (1998). QoS Impact on User Perception and Understanding of multimedia Video Clips, *Proc. of ACM Multimedia '98*, pp. 49-54, Bristol, UK.

Ghinea G., and Thomas J. P. (1999). Multimedia, Network Protocols and Users – Bridging the Gap, *Proc. of ACM Multimedia '99*, pp. 473-476, Orlando, Fl., USA.

Gulliver S.R., and Ghinea G. (2002). Impact of Captions on Deaf and Hearing Perception of Multimedia, *Proc. of ICME '02*, Lausanne, Switzerland.

Jayant N. S. (1990). High-quality coding of telephone speech and wideband audio *IEEE Communications Magazine*, Jan. 1990, 10-20.

Kawalek J. A. (1995). A User Perspective for QoS Management, *in Proceedings of the QoS Workshop aligned with the 3rd International Conference on Intelligence in Broadband Services and Network (IS&N 95)*, Crete, Greece, 16 September.

Watson A., and Sasse M.A. (2000). The Good, the Bad, and the Muffled: The Impact of Different Degradations on Internet Speech *Proc. of ACM Multimedia 2000*, pp. 269-302, Marina Del Rey, Ca., USA.

Royal National Institute for the Deaf. (2002). Fact sheet - Deaf and Hard of Hearing people, available at http://www.rnid.org.uk/, last accessed December 12.